Statistical Tests(Mann-Whitney)

from scipy.stats import mannwhitneyu

dataset.columns

Out[21]:  
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',  
       'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',  
       'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',  
       'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',  
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',  
       'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],  
      dtype='object')

a1=dataset.YearsAtCompany  
  
a2=dataset.YearsSinceLastPromotion  
  
stat,p=mannwhitneyu(a1,a2)  
  
print(stat,p)  
3672144.0 0.0

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted.

H0: There is no significant differences in the Years At Company and and Years Since Last Promotion

Ha: There is significant differences in the  Years At Company and Years SinceLast Promotion

Job Level vs Stock Option Level

a1=dataset.JobLevel  
  
a2=dataset.StockOptionLevel  
  
stat,p=mannwhitneyu(a1,a2)  
  
print(stat,p)  
3515436.0 0.0

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted.

H0: There is no significant differences in the Job Level and Stock Option Level

Ha: There is significant differences in the Job Level and Stock Option Level

Total Working Years vs Years Since Last Promotion

a1=dataset.TotalWorkingYears  
  
a2=dataset.YearsSinceLastPromotion  
  
stat,p=mannwhitneyu(a1,a2)  
  
print(stat,p)  
1907464.5 0.0

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted.

H0: There is no significant differences in the Total Working Years and and Years Since Last Promotion

Ha: There is significant differences in the Total Working Years and Years SinceLast Promotion

Percent Salary Hike vs Total Working Years

a1=dataset.PercentSalaryHike  
  
a2=dataset.TotalWorkingYears  
  
stat,p=mannwhitneyu(a1,a2)  
  
print(stat,p)  
4985332.5 0.0

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted.

H0: There is no significant differences in the Percent Salary Hike and Total Working Years

Ha: There is significant differences in the Percent Salary Hike and Total Working Years

Statistical Tests(Seperate T Test)

from scipy.stats import ttest\_ind  
  
dataset.columns  
Out[45]:  
Index(['Age', 'Attrition', 'BusinessTravel', 'Department', 'DistanceFromHome',  
       'Education', 'EducationField', 'EmployeeCount', 'EmployeeID', 'Gender',  
       'JobLevel', 'JobRole', 'MaritalStatus', 'MonthlyIncome',  
       'NumCompaniesWorked', 'Over18', 'PercentSalaryHike', 'StandardHours',  
       'StockOptionLevel', 'TotalWorkingYears', 'TrainingTimesLastYear',  
       'YearsAtCompany', 'YearsSinceLastPromotion', 'YearsWithCurrManager'],  
      dtype='object')  
  
a1=dataset.YearsAtCompany  
  
a2=dataset.YearsSinceLastPromotion  
  
stat,p=ttest\_ind(a1,a2)  
  
print(stat,p)  
46.25414163029768 0.0

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted

.    
H0: There is no significant differences in the Years At Company and Years Since Last Promotion

Ha: There is significant differences in the Years At Company and Years Since Last Promotion

from scipy.stats import pearsonr  
stats,p=pearsonr(dataset.DistanceFromHome,dataset.Age)  
print(stats,p)  
if(p<0.05):  
    print("reject null hypothesis")  
else:  
    print("accept null hypothesis")

Out[46]:

accept null hypothesis

Total Working Years vs Years Since Last Promotion

z1=dataset.TotalWorkingYears  
  
z2=dataset.YearsSinceLastPromotion  
  
stat,p=ttest\_ind(z1,z2)  
  
print(stat,p)

71.74859575019659 0.0

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted.    
H0: There is no significant differences in the Total Working Years and Year Since Last Promotion  
Ha: There is significant differences in the Total Working Years and Year Since Last Promotion

Job Level vs Stock Option Level

z1=dataset.JobLevel  
  
z2=dataset.StockOptionLevel  
  
stat,p=ttest\_ind(z1,z2)  
  
print(stat,p)

60.391685443496 0.0

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted.    
H0: There is no significant differences in the Job Level and Stock Option Level  
Ha: There is significant differences in the Job Level and Stock Option Level.

Percent Salary Hike vs Total Working Years

z1=dataset.PercentSalaryHike  
  
z2=dataset.TotalWorkingYears  
  
stat,p=ttest\_ind(z1,z2)  
  
print(stat,p)  
30.3706746654284 8.843533436260373e-193

As the P value is again 0.0, which is < than 0.05, the H0 is rejected and Ha is accepted.

H0: There is no significant differences in the Percent Salary Hike and Total Working Years.  
Ha: There is significant differences in the Percent Salary Hike and  Total Working Years.